

CLAIMS

What is claimed is:

1 1. A method for document classification comprising:
2 analyzing textual and graphical properties of an electronic document;
3 generating a classification of the document based on the textual and
4 graphical properties; and
5 storing the electronic document in a pre-existing document hierarchy
6 based on the classification.

1 2. The method defined in Claim 1 where the document hierarchy
2 comprises a hierarchy of documents mirroring in a similar fashion an
3 organization in a pre-existing memory storing documents.

1 3. The method defined in Claim 2 when the pre-existing memory
2 comprises a hard disk.

1 4. The method of claim 1, wherein analyzing textual properties of
2 an electronic document comprises:
3 determining characteristic words of the document;
4 determining a frequency for each characteristic word; and
5 building a frequency table based on the frequency associated with each
6 ~~characteristic word.~~

1 5. The method of claim 1, wherein analyzing graphical properties
2 of an electronic document comprises:
3 determining a point set corresponding to the electronic document,
4 wherein points of the point set correspond to end points of lines;
5 determining a density of points within the point set;
6 generating a document profile based, at least in part, on the density of
7 points within the point set.

1 6. The method of claim 1, wherein generating a classification of the
2 document based on the textual and graphical properties comprises combining
3 results from the textual and graphical analysis using a Borda Count.

1 7. The method defined in Claim 1 further comprising building the
2 pre-existing directory structure by building hierarchy of documents based on a
3 user's hard drive.

1 8. The method defined in Claim 1 further comprising building the
2 pre-existing directory structure by extracting graphical and text features from
3 documents in a directory-based memory to obtain a

5 determine a point set corresponding to the electronic document,
6 wherein points of the point set correspond to end points of lines;
7 determine a density of points within the point set;
8 generate a document profile based, at least in part, on the density of
9 points within the point set.

1 12. The machine-readable medium of claim 5, wherein the
2 sequences of instructions that cause the processor to generate a classification
3 of the document based on the textual and graphical properties further
4 comprises sequences of instructions that cause the processor to combine
5 results from the textual and graphical analysis using a Borda Count.

1 13. A method for document classification comprising:
2 analyzing documents in a pre-existing document directory structure to
3 determine an organization of the pre-existing document directory structure;
4 generating a mirror directory structure based on the pre-existing
5 document directory structure; and
6 placing a document in the mirror directory structure based on the
7 organization of the pre-existing document directory structure, results of
8 textual analysis of the document, and results of graphical analysis of the
9 ~~document.~~

4 determining a primary directory in the pre-existing document directory
5 structure in which the document should be placed based on the organization
6 of the pre-existing document directory structure; and

7 storing the document in a primary corresponding directory in the
8 mirror directory structure that corresponds to the primary directory in the
9 pre-existing document directory structure.

1 18. The method of claim 17 further comprising:

2 determining a secondary directory in the pre-existing document
3 directory in which the document should be placed based on the organization
4 of the pre-existing document directory structure; and

5 storing the document in a corresponding secondary directory in the
6 mirror directory structure that corresponds to the secondary directory in the
7 pre-existing document directory structure.

1 19. A computer-readable medium having stored thereon sequences
2 of instructions which, when executed by a processor, cause the processor to:

3 analyze a pre-existing document directory structure to determine an
4 organization of the pre-existing document directory structure;

5 generate a mirror directory structure based on the pre-existing
6 document directory structure; and

7 place a document in the mirror directory structure based on the
8 organization of the pre-existing document directory structure.

00000-0125100

1 20. The computer-readable medium of claim 19, wherein the
2 sequences of instructions that cause the processor to analyze a pre-existing
3 document directory structure to determine an organization of the pre-existing
4 document directory structure further comprise sequences of instructions that
5 cause the processor to:
6 recursively descending the pre-existing document directory structure;
7 generating a list of directories in the pre-existing document directory
8 structure;
9 examining files in directories of the pre-existing document directory
10 structure to determine content; and
11 examining the content of the files to determine the organization of the
12 directories in the pre-existing document directory structure.

1 21. The computer-readable medium of claim 19, wherein the
2 sequences of instructions that cause the processor to generate a mirror
3 directory structure based on the pre-existing document directory structure
4 further comprise sequences of instructions that cause the processor to
5 generate a document directory structure having a set of directories and
6 relationships equivalent to the pre-existing document directory structure.

1 22. The computer-readable medium of claim 19, wherein the
2 sequences of instructions that cause the processor to place a document in the

mirror directory structure based on the organization of the pre-existing document directory structure further comprise sequences of instructions that cause the processor to:

determine a primary directory in the pre-existing document directory structure in which the document should be placed based on the organization of the pre-existing document directory structure; and

store the document in a primary corresponding directory in the mirror directory structure that corresponds to the primary directory in the pre-existing document directory structure.

23. The computer-readable medium of claim 22 further comprising sequences of instructions that cause the processor to:

determine a secondary directory in the pre-existing document directory in which the document should be placed based on the organization of the pre-existing document directory structure; and

store the document in a corresponding secondary directory in the mirror directory structure that corresponds to the secondary directory in the pre-existing document directory structure.

24. An apparatus comprising:

means for analyzing a pre-existing document directory structure to determine an organization of the pre-existing document directory structure;

4 means for generating a mirror directory structure based on the pre-
5 existing document directory structure; and

6 means for placing a document in the mirror directory structure based
7 on the organization of the pre-existing document directory structure.

1 25. The apparatus of claim 24, wherein means for analyzing the pre-
2 existing document directory structure to determine the organization further
3 comprises:

4 means for recursively descending the pre-existing document directory
5 structure;

6 means for generating a list of directories in the pre-existing document
7 directory structure;

8 means for examining files in directories of the pre-existing document
9 directory structure to determine content; and

10 means for examining the content of the files to determine the
11 organization of the directories in the pre-existing document directory
12 structure.

1 26. The apparatus of claim 24, wherein means for generating a
2 mirror directory structure based on the pre-existing document directory
3 structure comprises means for generating a document directory structure
4 having a set of directories and relationships equivalent to the pre-existing
5 ~~document directory structure.~~

